

**Московский государственный университет
имени М.В. Ломоносова**

На правах рукописи



Арефьев Николай Викторович

**Методы построения и использования
компьютерных словарей сочетаемости для
синтаксических анализаторов русскоязычных текстов**

05.13.11 – математическое и программное обеспечение
вычислительных машин, комплексов и компьютерных сетей

Автореферат
диссертации на соискание ученой степени
кандидата физико-математических наук

Москва – 2012

Работа выполнена на кафедре алгоритмических языков факультета вычислительной математики и кибернетики Московского государственного университета имени М.В. Ломоносова.

Научный руководитель: доктор физико-математических наук, профессор Мальковский Михаил Георгиевич

Официальные оппоненты: Машечкин Игорь Валерьевич,
доктор физико-математических наук, профессор,
ВМК МГУ имени М.В. Ломоносова, профессор
кафедры АСВК, заведующий лабораторией

Фомичев Владимир Александрович,
доктор технических наук, доцент,
НИУ ВШЭ, факультет бизнес-информатики,
профессор кафедры инноваций и бизнеса в сфере
информационных технологий


Ведущая организация: Вычислительный центр РАН имени А.А. Дородницына

Защита состоится «28» сентября 2012 г. в 11.00 часов на заседании диссертационного совета Д 501.001.44 при Московском государственном университете имени М.В. Ломоносова по адресу: 119991, ГСП-1, Москва, Ленинские горы, МГУ, 2-ой учебный корпус, факультет вычислительной математики и кибернетики, аудитория 685. Желаящие присутствовать на заседании диссертационного совета должны сообщить об этом за 2 дня по тел. (495) 939-30-10 (для оформления заявки на пропуск).

С диссертацией можно ознакомиться в Фундаментальной библиотеке МГУ имени М.В. Ломоносова. С текстом автореферата можно ознакомиться на официальном сайте факультета ВМК МГУ <http://cs.msu.ru/> в разделе «Наука» – «Работа диссертационных советов» – «Д 501.001.44».

Автореферат разослан « 23 » августа 2012 г.

Председатель
диссертационного совета
член–корреспондент РАН
профессор



Королев Л.Н.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность работы. Несмотря на развитие технологий представления информации в формальном, доступном для компьютерной обработки виде, основной объем информации порождается, хранится и передается в виде текстов на естественном языке (ЕЯ). В связи с лавинообразным ростом количества информации в самых разных сферах человеческой деятельности возникает острая необходимость автоматического решения различных задач, связанных с обработкой текстов на ЕЯ, в том числе перевода с одного языка на другой, поиска информации в текстовых массивах, извлечения информации из текстов, реферирования и др.

Системы автоматической обработки текстов на ЕЯ (АОТ-системы), использующие методы поверхностного анализа (например, основанные на поиске ключевых слов) для решения перечисленных задач, в большинстве случаев не позволяют достичь результата, качество которого достаточно для практического применения. Причина кроется в необходимости учета не только слов, составляющих текст на ЕЯ, но и взаимосвязей между ними, не представленных в тексте в явном виде. Для выявления таких взаимосвязей требуется глубокий и полный анализ различных языковых явлений, представленных в тексте, и в первую очередь – выделение синтаксических отношений между словами текста (синтаксический анализ).

Для автоматического выделения синтаксических отношений требуется привлекать различную информацию о сочетаемости слов. Простейшим типом такой информации может служить формальное описание синтаксического поведения различных частей речи. Так, существительное сочетается с прилагательным при условии согласования их в роде, числе и падеже (*быстрый автомобиль*), но не сочетается с наречиями (**быстро автомобиль¹*). Помимо информации о сочетаемости частей речи в процессе автоматического синтаксического анализа необходимо привлекать и информацию об особенностях сочетаемости более узких классов слов, а также отдельных слов. Например, для того, чтобы правильно построить синтаксическую структуру фразы *одобрить поправки в закон* (*одобрить* → *поправки*, *поправки* → [*в закон*]) необходимо знать, что существительное *поправка* может управлять предложной группой с предлогом *в*, а глагол *одобрить* – не может (фраза *внести изменения в проект*, состоящая из других слов, принадлежащих тем же частям речи, имеет другую синтаксическую структуру: *внести* → *изменения*, *внести* → [*в проект*]). Отметим, что в лингвистике подобные ограничения называются морфосинтаксическими ограничениями на сочетаемость слов. Более сложная информация о сочетаемости требуется, чтобы построить правильные

¹ Здесь и далее символ ‘*’ указывает, что последующая фраза, нарушает правила языка.

синтаксические структуры фраз типа *привезти куртку на синтепоне* и *привезти куртку на машине*. В данном случае и глагол *привезти*, и существительное *куртка* сочетаются с предлогом *на*. Для выбора правильной структуры (для каждой из фраз своей) требуется привлечь формальные описания семантических (смысловых) ограничений на сочетаемость: *привезти* сочетается с классом слов, обозначающих средства передвижения (*на машине, на автобусе, на электричке*), а *куртка* – с названиями материалов (*на синтепоне, на меху, на пуху*). Наконец, в ряде случаев при синтаксическом анализе требуется информация о сочетаемости с конкретными лексемами (описания лексических ограничений на сочетаемость). Например, слова *проливной* и *сильный* близки по смыслу, но первое сочетается только со словом *дождь*, а второе – также с названиями других погодных явлений. Не обладая информацией об индивидуальной сочетаемости слов, синтаксический анализатор во фразе типа *проливные дожди и заморозки* может ошибочно отнести прилагательное *проливные* не к существительному *дожди*, а к группе однородных членов *дожди и заморозки* (по аналогии с фразой *сильные дожди и заморозки*, для которой такой вариант допустим).

Таким образом, для качественного решения задачи автоматического синтаксического анализа необходимо подробное и полное описание сочетаемостных ограничений перечисленных типов (морфо-синтаксических, семантических и лексических). Как показал обзор существующих в свободном доступе лингвистических описаний сочетаемости слов русского языка (словарей сочетаемости, комбинаторных словарей), данные источники информации о сочетаемости обладают существенными недостатками. Во-первых, большинство таких источников рассчитано на пользователя-человека, поэтому зачастую авторы вместо того, чтобы приводить формальное и последовательное описание сочетаемости некоторого слова, ограничиваются рядом примеров и ссылок на аналогичные слова, апеллируя к интуиции пользователя словаря. Во-вторых, доступные словари (в особенности те, которые формализованы в достаточной для практического применения степени) покрывают лишь небольшую часть лексики русского языка. В-третьих, в большинстве словарей сведения о семантических ограничениях на сочетаемость либо не приводятся вовсе, либо не формализованы в достаточной мере. Основной причиной перечисленных недостатков является чрезвычайно высокая трудоемкость ручного формирования описаний сочетаемости, носящих комбинаторный характер (по сути, требуется описать множество пар, или даже n -ок слов, способных образовывать допустимые словосочетания). Вследствие этих недостатков в задачах автоматической обработки текстов существующие ресурсы могут использоваться только как дополнительный источник информации о сочетаемости, обладающий весьма высокой точностью (словарные статьи

формируются компетентными экспертами-лингвистами и заслуживают доверия), но очень низкой полнотой.

Альтернативой использованию лингвистических описаний сочетаемости является автоматический сбор статистики совместной встречаемости слов на большой текстовой коллекции и формирование статистического описания сочетаемости. При этом имеет смысл использовать неразмеченные (т.е. не обработанные экспертами) тексты, поскольку создание достаточной по объему размеченной коллекции является очень сложной и трудоемкой задачей. Такой подход позволяет свести к минимуму объем требуемого ручного труда, а также обеспечить довольно полный охват лексики. Однако простая статистика совместной встречаемости слов не дает всей необходимой информации о сочетаемости. Это связано с проблемой разреженности данных о совместной встречаемости, извлеченных из коллекции текстов на ЕЯ: лишь небольшая часть сочетающихся между собой слов реально встретятся вместе в коллекции. Свойство разреженности является фундаментальным для текстов на ЕЯ, поэтому решить данную проблему невозможно ни увеличением объема, ни изменением состава текстовой коллекции. Особенно остро проблема разреженности встает в тех случаях, когда описываемое слово сочетается со всеми словами, принадлежащими одному или нескольким достаточно широким семантическим классам. Например, слово *пирог* сочетается с названиями практически любых неожиданных пищевых продуктов (*пирог с черникой, творогом, рыбой, вареньем, маком...*), однако подобрать коллекцию текстов, в которую войдут все такие словосочетания невозможно, а описать только морфологические и синтаксические характеристики сочетающихся слов – означает допустить словосочетания типа *пирог с идеей*. Для решения данной проблемы необходимо использовать механизмы обобщения, которые на основе встречаемости слова с представителями определенных семантических классов делают вывод о сочетаемости со всеми словами, относящимися к этим классам. При этом встает вопрос об источнике информации о семантических классах и принадлежащих им словах. Большинство существующих методов (ориентированных в основном на обработку англоязычных текстов) в качестве такого источника используют лексическую онтологию WordNet², являющуюся довольно качественным лингвистическим ресурсом с широким охватом лексики английского языка. Однако, для многих других языков (в том числе для русского) аналоги WordNet, обладающие сопоставимой полнотой и качеством описания, недоступны и создание их упирается в те же проблемы, что и создание словарей сочетаемости. Альтернативой опять же является автоматическое извлечение информации о семантических классах из неразмеченных текстовых коллекций.

² <http://wordnet.princeton.edu/>

Существующие методы автоматического формирования описаний сочетаемости и семантических классов не дают результатов, достаточно качественных для того, чтобы можно было полностью исключить ручной труд эксперта. Они, однако, способны дать первое приближение к необходимому описанию и обеспечить эксперта статистическими данными для дальнейшего улучшения этого описания.

Таким образом, актуальным является создание методов автоматизированного формирования описаний сочетаемости, позволяющих извлекать информацию о сочетаемости из неразмеченных текстовых коллекций, обобщать ее и представлять в таком виде, в котором эксперты могут эффективно работать с ней. Другой актуальной проблемой является учет сформированных таким образом, а также содержащихся в существующих словарях, описаний сочетаемости для улучшения качества и повышения эффективности автоматического синтаксического анализа.

Цель работы. Целью диссертационной работы является исследование существующих и разработка новых методов построения компьютерных словарей сочетаемости и использования этих словарей для повышения качества работы автоматических синтаксических анализаторов русскоязычных текстов.

Разрабатываемые методы должны поддерживать автоматизированный подход к формированию словарей, при котором в процессе формирования используются как алгоритмы автоматического извлечения информации о сочетаемости из текстовых коллекций, так и данные, полученные от экспертов или из составленных вручную лингвистических ресурсов. При этом участие экспертов должно быть максимально эффективным, для чего необходимо создать соответствующие инструментальные средства развития и сопровождения словарей сочетаемости.

Достижение поставленной цели предполагает решение следующих задач:

- 1) исследование ограничений на сочетаемость, которые необходимо учитывать в процессе синтаксического анализа, и разработка структур данных, позволяющих хранить описания таких ограничений;
- 2) разработка методов и программных средств, позволяющих учитывать сочетаемостные ограничения в процессе синтаксического анализа;
- 3) разработка методов и программных средств автоматизированного формирования описаний сочетаемостных ограничений, их развития и сопровождения.

Работа выполнялась на кафедре алгоритмических языков факультета ВМК МГУ в рамках Госбюджетных НИР по приоритетному направлению "Программное и математическое обеспечение эффективного решения актуальных задач на современных вычислительных системах", проводимых под

руководством профессора, д.ф.-м.н. М.Г. Мальковского и подразумевающих, в числе прочего, создание системы автоматического синтаксического анализа русскоязычных текстов Treeton. В системе Treeton сочетаемость слов изначально не учитывалась, что негативно сказывалось на качестве ее работы. В связи с этим актуальной стала задача интеграции в данную систему программных средств формирования и учета описаний сочетаемости в процессе синтаксического анализа.

Методы исследования. Для решения поставленных задач использовались методы теории вероятности и математической статистики, машинного обучения и интеллектуального анализа данных, объектно-ориентированного анализа и проектирования.

Основные результаты. Основные результаты работы заключаются в следующем:

1. Исследованы классы ограничений (на сочетаемость слов), которые целесообразно учитывать при синтаксическом анализе, и для них разработаны структуры данных, позволяющие комбинировать статистический и онтологический подходы к описанию ограничений.
2. Разработаны методы, алгоритмы и технология автоматизированного формирования компьютерных словарей сочетаемости, их развития и сопровождения, а также поддерживающие предложенную технологию инструментальные средства.
3. Предложены методы учета словарной информации о сочетаемости, позволяющие повысить эффективность синтаксического анализа.

Научная новизна. Предложен новый комплексный подход к проблеме повышения качества работы синтаксических анализаторов русскоязычных текстов за счет учета морфо-синтаксических, лексических и семантических ограничений на сочетаемость слов анализируемого текста. В рамках данного подхода информация о сочетаемости извлекается как из лингвистических источников, так и из корпусов текстов. Извлеченная информация подвергается дополнительной обработке с применением методов интеллектуального анализа данных для устранения разреженности. Информация о сочетаемости интегрирована в специальном компьютерном словаре сочетаемости, который может формироваться как полностью автоматически, так и полуавтоматически (с участием эксперта). Формирование и поддержка словаря обеспечены необходимыми инструментальными средствами. Такой подход позволяет компенсировать и нехватку адекватных лингвистических описаний сочетаемости, и недостаточную репрезентативность обучающей выборки (исходных корпусов текстов).

Практическая значимость. Разработанные методы формирования и использования компьютерных словарей сочетаемости могут применяться для улучшения качества работы различных приложений, решающих задачи автоматической обработки текстов на ЕЯ. Предложенные методы легли в основу подсистемы работы со словарями сочетаемости автоматического синтаксического анализатора русскоязычных текстов Treeton. Данная подсистема позволила улучшить качество работы синтаксического анализатора.

Также разработанные программные средства могут служить инструментом лингвистических исследований в области лексикографии при составлении словарей (в первую очередь словарей сочетаемости) для получения различной статистической информации о сочетаемости слов в корпусе.

Апробация работы. Полученные результаты докладывались и обсуждались на следующих конференциях и семинарах:

- международная конференция по компьютерной лингвистике «Диалог» (2010 г.);
- научная конференция МГУ «Тихоновские чтения» (2011 г.);
- международные научно-практические конференции SWorld (2008, 2009, 2011, 2012 гг.);
- научно-исследовательские семинары МГУ им. М.В. Ломоносова, ВЦ РАН, Института Русского языка РАН, Высшей школы экономики.

Публикации. Основные результаты диссертации отражены в десяти публикациях: в научном журнале из перечня ВАК [1], в двух тематических сборниках [2,3], а также в других изданиях [4-10]. В работах, выполненных в соавторстве с М.Г. Мальковским, М.Г. Мальковскому принадлежит постановка задачи и общее руководство исследованием, автору – решение задачи, включающее разработку методов и программных средств. В совместной с А.С. Старостиним работе [7] А.С. Старостину принадлежат результаты, связанные с разработкой формализма тринотаций для описания синтаксических структур и базового алгоритма автоматического синтаксического анализа, не учитывающего в процессе работы ограничения на сочетаемость слов; автору принадлежит разработка механизма учета ограничений на сочетаемость в процессе синтаксического анализа. В совместной с И.А. Булгаковым работе [3] автором была разработана и реализована инструментальная среда оценки качества и сопровождения синтаксического анализатора, в которую затем был интегрирован предложенный И.А. Булгаковым алгоритм оценки качества синтаксических анализаторов, использующих формализм тринотаций.

Структура и объем диссертации. Диссертация состоит из введения, пяти глав, заключения, списка литературы и приложений. Объем диссертации – 155

страниц, объем приложений – 33 страницы. Список литературы содержит 76 наименований.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во введении обоснована актуальность диссертационной работы, сформулированы цель и задачи исследования.

В первой главе приводится описание проблемы, решению которой посвящена диссертационная работа; рассматриваются существующие подходы к решению этой проблемы, присущие им достоинства и недостатки.

Одним из важнейших свойств ЕЯ является омонимия: знаки ЕЯ (морфемы, слова, словосочетания и предложения) допускают неоднозначную интерпретацию. В процессе синтаксического анализа приходится иметь дело с синтаксической омонимией – возможностью построения синтаксическим анализатором для одного предложения нескольких синтаксических структур. Причиной может служить как реальная неоднозначность анализируемого предложения (*историей математики интересуются: историей→математики* или *интересуются→математики*), так и несовершенство используемых при синтаксическом анализе алгоритмов и ресурсов, из-за которого система не может отбросить неверные синтаксические структуры (*съесть пирог с удовольствием: съесть→[с удовольствием]* или **пирог→[с удовольствием]*). В последнем случае омонимия является неоправданной с точки зрения ЕЯ и приводит к ухудшению качества (на выходе анализатора появляются неправильные варианты анализа предложения) и уменьшению скорости работы (на построение неправильных вариантов тратятся временные ресурсы) синтаксического анализатора. В разделе 1.1 рассмотрен алгоритм работы синтаксического анализатора Treeton. Выделены типы омонимии, возникающие в том случае, когда в процессе анализа не учитываются ограничения на сочетаемость слов. Показано, какая информация о сочетаемости необходима для того, чтобы разрешить неоправданную с точки зрения ЕЯ омонимию (т.е. выбрать из построенных синтаксических структур правильную).

Алгоритм работы синтаксического анализатора системы Treeton основан на идее перебора под управлением эвристической функции. На входе синтаксический анализатор получает набор морфологических интерпретаций слов анализируемого предложения (помеченных вершин будущей синтаксической структуры, соответствующих словам предложения). На каждом шаге переборного процесса применяется одно из заложенных в системе синтаксических правил. В результате применения правила либо порождается новая вершина, соответствующая синтаксической группе (например, группе однородных членов [*яблоки и груши*] в словосочетании *спелые яблоки и груши*), либо одна из вершин связывается с другой вершиной синтаксической связью, то

есть дугой, помеченной типом связи (например, *яблоки* →_{modif} *спелые* или [*яблоки и груши*] →_{modif} *спелые*). Если в результате применения правила получается структура, содержащая все слова анализируемого предложения (итоговая структура), она попадает в список результатов анализа. В противном случае получившаяся промежуточная структура участвует в дальнейшем переборе.

Явление омонимии (неоднозначности анализа), присущее естественному языку, приводит к комбинаторному взрыву при слепом поиске возможных вариантов синтаксической структуры предложения. Для борьбы с этой проблемой в синтаксическом анализаторе Treeton используются эвристики, делающие переборный процесс направленным. Любая промежуточная структура, полученная в процессе перебора, оценивается – ей приписывается вектор (называемый штрафным вектором), каждый из компонентов которого вычисляется соответствующей штрафной функцией. Каждая штрафная функция проверяет соответствие промежуточной структуры определенным языковым нормам и возвращает неотрицательное действительное число, которое тем больше, чем серьезнее структура нарушает данные нормы. Евклидова норма штрафного вектора называется штрафом данной структуры. Штраф влияет на выбор промежуточных структур, к которым будут применяться синтаксические правила на очередном шаге перебора (в первую очередь используются наименее штрафованные структуры), а также на положение итоговых структур в списке результатов анализа (структуры в данном списке упорядочены по возрастанию их штрафов). Результаты синтаксического анализа считаются правильными, если первой в списке результатов (наименее оштрафованной) является правильная синтаксическая структура анализируемого предложения. Кроме того, для эффективной работы анализатора требуется, чтобы неправильные синтаксические структуры были сильно оштрафованы на ранних этапах их обработки, а правильные, напротив, получили нулевые или небольшие штрафы. Таким образом, встает вопрос о выборе системы штрафных функций, обеспечивающих правильную и эффективную работу синтаксического анализатора.

В ранних версиях синтаксического анализатора использовались штрафные функции, учитывающие только топологические свойства оцениваемых структур: проективность, число одинаково помеченных дуг, исходящих из одной вершины и т.п. Хотя такие штрафные функции позволили существенно сократить перебор, они не дали удовлетворительного решения проблемы омонимии – число итоговых структур, имеющих небольшие штрафы, измерялось десятками и сотнями, и правильные структуры далеко не всегда оказывались первыми в списке результатов анализа. Как показано в первой главе, существуют широкие

классы русскоязычных предложений, правильный анализ которых возможен в том и только том случае, когда система штрафных функций при оценке структуры учитывает сочетаемость слов, связанных в рамках этой структуры синтаксическими связями. Таким образом, возникает задача оценки сочетаемости пары слов w_1 и w_2 , связанных синтаксической связью типа r (в дальнейшем такая конструкция обозначается как $w_1 \rightarrow_r w_2$).

В разделе 1.2 рассмотрены существующие подходы к описанию сочетаемости и ее учету в процессе синтаксического анализа. Вначале на примере системы ЭТАП³ рассмотрены достоинства и недостатки подхода, при котором сочетаемость описывается вручную экспертами-лингвистами. К достоинствам можно отнести возможность формирования таким способом более точных и исчерпывающих описаний, нежели автоматическими методами. Недостатками является чрезвычайно высокие трудоемкость и стоимость создания словарей сочетаемости вручную. К тому же лингвисты лучше справляются с описанием качественных, но не количественных характеристик сочетаемости, причем созданные ими описания зачастую обладают низкой степенью формализованности, что затрудняет их использование в АОР-системах. В связи с этим присутствующие в свободном доступе словари сочетаемости покрывают лишь небольшую часть лексики русского языка, требуют дополнительной работы по формализации содержащихся в них описаний и могут использоваться скорее не как основной, а как дополнительный источник информации о сочетаемости. Как показывает опыт разработки словарей системы ЭТАП (отсутствующих в свободном доступе) для достижения приемлемых результатов при ручном формировании требуется несколько десятилетий работы большого коллектива высококвалифицированных специалистов. В рамках системы ЭТАП также проводились эксперименты по автоматическому формированию словаря сочетаемости на основе синтаксически размеченной экспертами коллекции текстов. Выяснилось, что ручная синтаксическая разметка коллекции текстов достаточного объема является не менее трудоемкой задачей, нежели ручное формирование словаря сочетаемости, поэтому проблем высокой стоимости и трудоемкости создания словаря таким способом избежать не удастся.

Использование автоматических методов извлечения информации о сочетаемости из неразмеченных коллекций текстов (на основании статистики совместной встречаемости слов) представляется намного более реалистичным подходом, с которым, однако, также связан ряд трудностей. Во-первых, даже

³ Jurij Apresian, Igor Boguslavsky, Leonid Iomdin, Alexander Lazursky, Vladimir Sannikov, Victor Sizov, Leonid Tsinman. ETAP-3 Linguistic Processor: a Full-Fledged NLP Implementation of the MTT // MTT 2003. First International Conference on Meaning-Text Theory. Paris: Ecole Normale Supérieure, 2003. – С. 279–288.

при использовании очень больших текстовых коллекций полученные статистические данные оказываются неполны. Это связано со свойством разреженности, неотъемлемо присущим коллекциям текстов на ЕЯ. Во-вторых, зачастую возникают проблемы с обработкой сложных синтаксических конструкций, характерных для большинства предложений в текстах на ЕЯ. Если при сборе статистики обрабатывать сложные конструкции, неизбежно растет (в силу недостатков существующих методов) число ошибок в получаемых статистических данных. Если же такие конструкции игнорировать, радикально уменьшается объем обучающей выборки. В целом статистические методы в чистом виде годятся лишь для разрешения некоторых достаточно простых типов омонимии.

Для того чтобы собранные на корпусе статистические данные оказались полезны для решения практических задач, необходима интеллектуальная обработка этих данных. Примером такой обработки может служить построение на основе отдельных слов, встретившихся с данным словом, набора семантических классов, сочетающихся с данным словом (обобщение). Для эффективного получения качественного результата при такой обработке требуется привлекать как методы машинного обучения (такие как кластеризация, классификация), так и знания экспертов. Методы машинного обучения позволяют эксперту иметь дело не с огромными списками слов и таблицами частот их совместной встречаемости, а с некоторыми классами слов, работа с которыми удобнее и продуктивнее. Знания эксперта позволяют устранять ошибки, неизбежно возникающие на различных этапах обработки данных из-за чрезвычайно высокой сложности естественного языка.

Во второй главе описана структура компьютерного словаря сочетаемости (подсистемы, предназначенной для хранения информации о сочетаемости слов).

В разделе 2.1 сформулированы требования к компьютерному словарю сочетаемости, вытекающие, с одной стороны, из лингвистических реалий, для моделирования которых он предназначен, а с другой стороны, из потребностей синтаксического анализатора в информации о сочетаемости слов, которые словарь призван удовлетворить. В лингвистике выделяют три базовых типа ограничений на сочетаемость: морфо-синтаксические, семантические и лексические. Морфо-синтаксическими ограничениями на сочетаемость слова w называются ограничения на часть речи и грамматические характеристики слов w' , сочетающихся с w . Семантические ограничения на сочетаемость имеют место в тех случаях, когда слово w сочетается с любыми словами w' , принадлежащими определенным семантическим классам. Лексические ограничения на сочетаемость проявляются в тех случаях, когда слово w сочетается только с определенными словами, сгруппировать которые в

семантические классы затруднительно. Как показано в 2.1, ограничения всех трех типов необходимо учитывать в процессе синтаксического анализа, а значит, и хранить в словаре сочетаемости. Рассмотрим особенности перечисленных типов ограничений на сочетаемость, которые отражаются на структуре компьютерного словаря сочетаемости.

Слово может накладывать ограничения не только на слова, связанные с ним непосредственно синтаксической связью. Простейшим примером могут служить предложные группы – хозяин предложной группы управляет зависимым существительным не непосредственно, а через предлог (*оказаться*→*в*→*дамках*). Для подобных случаев автором было введено понятие метасвязи: считается, что между хозяином предложной группы и зависимым от предлога существительным существует метасвязь (связь, не являющаяся синтаксической), а предлог входит в название этой связи (*оказаться*→_{circumst_v_pp} *дамках*). За счет введения метасвязей становится возможным единообразно описывать самые разные ограничения на сочетаемость.

Морфо-синтаксические ограничения должны указываться для слова, а лексические и семантические – для пары <слово, тип связи> (под связью понимается либо синтаксическая связь, либо метасвязь). Морфо-синтаксические ограничения описываются как набор возможных для данного слова типов связей. Лексические ограничения представляются списком слов, которые могут находиться на другом конце связи, а семантические ограничения – списком возможных семантических классов таких слов.

Весьма важным является вопрос о представлении семантических классов в словаре. Простейшим подходом, используемом в частности в системе ЭТАП, является представление семантического класса в виде списка слов, которому поставлен во взаимно-однозначное соответствие некоторый уникальный идентификатор. Данный идентификатор указывается при описании семантической сочетаемости вместо соответствующего списка слов. У такого подхода имеется существенный недостаток: в случае наличия двух вариантов синтаксического анализа, каждый из которых удовлетворяет семантическим ограничениям, либо ни один из которых не удовлетворяет им, выбрать из этих вариантов лучший на основе семантических соображений не удастся. Подобная ситуация может возникнуть в случае неполноты (для некоторого семантического класса приведены не все слова, к нему относящиеся; для некоторого слова перечислены не все семантические классы, с которыми он сочетается), либо неточности (для некоторого слова указан слишком широкий семантический класс, включающий слова, с которыми данное слово не сочетается) словарной информации. Заметим, что от неполноты и неточности при составлении словаря на практике невозможно полностью избавиться. Также подобное представление

семантических классов становится проблемой в пограничных случаях, когда существуют аргументы как за, так и против включения некоторого слова в семантический класс. Для решения указанных проблем автором было введено представление семантического класса как нечеткого множества слов; в этом случае отношение соответствия слова семантическим ограничениям на сочетаемость становится нечетким отношением.

Можно выделить два типа информации о сочетаемости слов. Информация первого типа (назовем ее статистической информацией о сочетаемости) представляет собой численные оценки сочетаемости, которые могут быть получены автоматически в результате сбора статистики совместной встречаемости слов в текстовой коллекции. Поскольку такая информация формируется автоматически, ею можно снабдить все слова в словаре сочетаемости (или, по крайней мере, те, которые встречаются в коллекции достаточное для получения достоверных статистических данных число раз). Однако статистическая информация неизбежно является зашумленной (содержит ошибки), что вызвано как ошибками обработки текстовой коллекции, так и ошибками и отклонениями от норм языка в самой коллекции. Информация второго типа (будем называть ее онтологической информацией о сочетаемости) представляет собой данные о сочетаемости, содержащиеся в различного рода лингвистических источниках, составляемых экспертами. Такая информация обычно носит нечисловой характер. Она содержит мало ошибок по сравнению с данными, полученными автоматически, но при этом имеется лишь для ограниченного множества слов (что связано с высокой трудоемкостью составления описаний сочетаемости вручную). В словаре сочетаемости хранится информация обоих типов. Хранение статистической информации дает высокую полноту описания сочетаемости. Хранение онтологической информации позволяет обеспечить высокую точность описания сочетаемости для ограниченного множества слов (например, для терминов некоторой предметной области, точность описания которой часто критична для качества анализа текстов этой предметной области).

В разделе 2.2 предлагается система классов на языке Java, реализующая словарную подсистему, удовлетворяющую сформулированным в 2.1 требованиям. Вначале описаны основные операции, которые поддерживает словарная подсистема, затем приводятся структуры данных, обеспечивающие эффективное хранение информации о сочетаемости и выполнение данных операций. Основной структурой данных, содержащей информацию о сочетаемости, является тензор сочетаемости – трехмерный массив, два измерения которого соответствуют словам и семантическим классам, а третье – типам связей. В ячейке тензора $\langle from, rel, to \rangle$ хранится информация о

сочетаемости слов или семантических классов *from* и *to* по связи типа *rel*. Информация о сочетаемости представлена в виде набора атрибутов и их значений. Среди атрибутов есть атрибуты, представляющие как статистическую информацию о сочетаемости (частота конструкции $from \rightarrow_{rel} to$ в корпусе, условная вероятность $P(to | from, rel)$ и др.), так и онтологическую (наличие типа связи *rel* в модели управления слова *from*, экспертная оценка допустимости конструкции $from \rightarrow_{rel} to$ и др.). На основе значений всех хранимых атрибутов в процессе синтаксического анализа вычисляется итоговая оценка соответствия конструкции $from \rightarrow_{rel} to$ ограничениям сочетаемости.

Важную роль в описании семантической сочетаемости играет понятие контекстного вектора. Контекстный вектор слова или семантического класса *from* представляет собой вектор, компоненты которого соответствуют ячейкам слоя $\langle from, *, * \rangle$ тензора сочетаемости, а значениями компонент являются значения атрибута $prmi^4$ (меры неслучайности) соответствующих ячеек. С помощью контекстных векторов вводится расстояние между словами, а также мера принадлежности слова семантическому классу. Расстояние между словами w_1 и w_2 определяется через косинус угла между их контекстными векторами $v(w_1)$ и $v(w_2)$: $\rho(w_1, w_2) = 1.0 - \cos(v(w_1), v(w_2))$. Мера принадлежности слова w классу C определяется похожим образом: $\mu_C(w) = \cos(v(w), v(C))$.

Тензор сочетаемости чрезвычайно разрежен (большинство ячеек тензора пусты), что позволяет применять стандартные методы хранения разреженных массивов для эффективного хранения тензора. Также с учетом разреженности был реализован эффективный алгоритм вычисления косинуса угла между контекстными векторами.

В третьей главе предлагается подход к решению проблемы автоматизированного формирования словаря сочетаемости (то есть заполнения структур данных, описанных во второй главе, информацией о сочетаемости конкретных слов). В разделе 3.1 описан процесс автоматического извлечения статистической информации о сочетаемости из коллекции текстов, а в разделе 3.2 решается задача импорта онтологической информации о сочетаемости из лингвистических источников в словарь.

Процесс извлечения статистической информации о сочетаемости состоит из четырех этапов: формирование корпуса примеров, его морфологический анализ, частичный синтаксический анализ и статистическая обработка. Обозначим через w_1 слово, информацию о сочетаемости которого необходимо извлечь. На первом этапе решается задача формирования на основе исходной текстовой коллекции оптимального по объему и содержанию корпуса предложений, из которого за

⁴ Gerlof Bouma. Normalized (Pointwise) Mutual Information in Collocation Extraction // From Form to Meaning: Processing Texts Automatically, Proceedings of the Biennial GSCL Conference 2009. Tübingen : Gunter Narr Verlag, 2009. – P. 31-40.

приемлемое время можно извлечь достоверную статистическую информацию о сочетаемости слова w_1 . В качестве исходной текстовой коллекции был выбран Национальный Корпус Русского Языка⁵ (далее НКРЯ), являющийся одной из наиболее представительных коллекций текстов на русском языке. Корпус примеров формируется на основе запроса, задаваемого на разработанном автором языке запросов к НКРЯ. Запрос состоит из шаблона предложений, которые необходимо включить в корпус примеров, и дополнительных ограничений на количество и длину таких предложений. Результат запроса сохраняется в локальной файловой системе на случай повторного обращения. На втором этапе производится морфологический анализ сформированного корпуса примеров. Используется доработанный морфологический анализатор системы Treetop, в который автором были добавлены механизмы частичного снятия морфологической омонимии и морфологического анализа отсутствующих в словаре слов. На третьем этапе производится частичный синтаксический анализ корпуса. Синтаксический анализатор системы Treetop решает задачу полного синтаксического анализа, то есть построения дерева (или нескольких деревьев для омонимичных предложений), включающего все слова исходного предложения. Если такое дерево построить не удастся (например, из-за наличия в предложении синтаксических явлений, не описанных в системе правил), то результатом работы анализатора будет пустое множество синтаксических структур. Это позволяет обеспечить высокую точность анализа отдельных предложений, но приводит к низкой полноте при анализе корпуса. Также алгоритм полного синтаксического анализа является переборным, что приводит к ощутимому падению скорости анализа на длинных предложениях. Для извлечения статистической информации не требуется полный синтаксический анализ корпуса – достаточно построения синтаксических связей для того слова, информация о сочетаемости которого извлекается. Кроме того не требуется столь высокая точность как при анализе одного предложения – ошибки анализа отдельных предложений корпуса не вносят существенный вклад в собираемую статистику (если они не носят систематический характер). В то же время требуется приемлемая скорость анализа большого числа предложений. Для удовлетворения данных требований автором был разработан инструмент частичного синтаксического анализа, который для входного предложения строит набор связей $w_1 \rightarrow_r w_2$, где w_1, w_2 – начальные формы слов предложения (построенные морфологическим анализатором), а r – тип связи. При этом на структуры, которые строятся в результате частичного синтаксического анализа (в отличие от полного) не накладываются требования древовидности и покрытия всех слов исходного предложения, что позволяет существенно улучшить

⁵ <http://www.ruscorpora.ru/>

полноту анализа, упростить алгоритм и ускорить его работу. На последнем, четвертом этапе результаты частичного синтаксического анализа всех предложений корпуса подвергаются статистической обработке. В частности, вычисляются и заносятся в соответствующие ячейки тензора частоты $f(w_1, r, w_2)$ связей слова w_1 , а также значения меры $ppmi$ их неслучайности, являющиеся компонентами контекстного вектора $v(w)$:

$$\langle w_1, r, w_2 \rangle_{ppmi} = \max(0, \log \frac{\hat{P}(w_1, r, w_2)}{\hat{P}(w_1)\hat{P}(w_2)}).$$

Смысл этой меры заключается в следующем: чем больше появление слова w_1 увеличивает апостериорную вероятность появления связанного с ним связью r слова w_2 , тем больше величина меры неслучайности. Мера неслучайности позволяет оценить, насколько та или иная связь специфична именно для данного слова, и придать более специфичным связям больший вес при вычислении расстояния между словами.

В разделе 3.2 описан процесс импорта в словарь сочетаемости онтологической информации о сочетаемости из двух лингвистических источников: базы данных моделей управления⁶ и толкового словаря⁷ (отметим, что метод применим и ко многим другим толковым словарям, имеющим схожую структуру словарной статьи). Первый дает описания морфо-синтаксической сочетаемости слов, а второй – информацию о принадлежности слов семантическим классам (такая информация помогает оценить семантическую сочетаемость слов, для которых построить контекстные векторы невозможно из-за недостаточной репрезентативности корпуса примеров).

В четвертой главе предложены методы обобщения статистической информации о сочетаемости, полученной в результате автоматической обработки текстовой коллекции. В разделе 4.1 показано, что такой информации присущи недостатки – зашумленность (часть связей, которые строит частичный синтаксический анализатор, представляет собой примеры неправильной или нетипичной сочетаемости слов) и разреженность (лишь небольшая часть сочетающихся друг с другом слов встречается вместе в конкретной текстовой коллекции, даже если эта коллекция очень объемна). Зашумленность можно существенно снизить, если отбросить низкочастотные связи слова, среди которых много ошибочных. При этом, однако, еще острее встает проблема разреженности – в ряде экспериментов даже при отбрасывании лишь тех связей, которые встретились только один раз, терялась практически половина

⁶ Перцов Н.В., Старостин С.А. О синтаксическом процессоре, работающем на ограниченном объеме лингвистических средств // Труды международной конференции Диалог'1999. Т.2. Таруса, 1999. – С. 224-230.

⁷ Ожегов С.И., Шведова Н.Ю. Толковый словарь русского языка: 80 000 слов и фразеологических выражений. М.: Азъ, 1993. – 955 с.

правильных связей слова. Для решения данных проблем и предназначены предложенные автором методы обобщения, позволяющие устанавливать правильность связи независимо от ее частоты (и вообще наличия) в корпусе. Основная идея обобщения заключается в переходе от встретившихся в корпусе связей к некоторым классам допустимых, но необязательно встретившихся в корпусе, связей.

В разделе 4.2 обсуждаются возможные подходы к обобщению. Поскольку связь представляет собой тройку, состоящую из двух слов и типа связи, можно предложить два варианта обобщения: обобщение по словам и по типу связи. Обобщение по типу связи реализуется с помощью метасвязей. Например, для словосочетаний *крепкий чай* и *чай был крепким* порождается одна и та же метасвязь *чай*→*крепкий*. Это позволяет автоматически делать вывод о допустимости обоих словосочетаний, даже если в корпусе встретилось только одно из них. Обобщение по словам реализуется переходом от конкретных слов к семантическим классам. Метод, позволяющий осуществить такой переход, основан на представлении слов в виде контекстных векторов и применении методов интеллектуального анализа данных к последним. В разделе 4.3 обсуждаются различные способы построения контекстных векторов слов и вычисления расстояния между ними, выбирается способ, при котором расстояние между контекстными векторами отражает смысловую близость соответствующих этим векторам слов. В разделе 4.4 описан алгоритм разбиения заданного множества слов на группы (кластеры) близких по смыслу слов, в основе которого лежит процедура агломеративной иерархической кластеризации множества контекстных векторов этих слов. В разделе 4.5 приводится описание того, как предложенный алгоритм используется для автоматического формирования описаний семантической сочетаемости. Если на вход алгоритма подается множество слов, семантически сочетающихся с некоторым словом w , то набор кластеров на выходе алгоритма представляет собой некоторое приближение набора тех семантических классов, с которыми сочетается w . Описанием семантической сочетаемости w служит набор центроидов полученных кластеров (т.е. векторов, представляющих собой среднее арифметическое контекстных векторов слов каждого кластера). Для произвольного слова w' можно вычислить расстояние от его контекстного вектора до каждого из центроидов; минимальное из этих расстояний является мерой семантической сочетаемости слова w' со словом w . Таким образом, происходит обобщение по словам – слово w' может быть признано семантически сочетающимся с w , даже если оно ни разу не встретилось с ним в корпусе.

Качество такого обобщения очень сильно зависит от того набора слов, который подается на вход алгоритма кластеризации. Если в этом наборе есть

слова, сочетающиеся с w лексически или вообще не сочетающиеся с ним, то слова, близкие к ним по смыслу, будут признаны (ошибочно) также сочетающимися с w (строить обобщение можно только на основе слов, семантически сочетающихся с данным). Также проблемы могут возникнуть, если среди кластеризуемых слов мало представителей некоторого семантического класса. В этом случае среднее арифметическое контекстных векторов этих представителей может неадекватно описывать данный семантический класс. Таким образом, возникает задача выбора среди слов, которые встретились с данным, тех слов, которые сочетаются с ним семантически (выявления семантической сочетаемости). Предложенный автором метод решения этой задачи описан в разделе 4.6. Вначале из множества всех связей типа r слова w удаляются связи с низкой частотой $f(w, r, w')$. Это позволяет отсеять большую часть ошибочных связей, но вместе с ними теряется и существенный процент правильных связей. Назовем множество слов w' , входящих в полученное множество связей типа r , базовым множеством. На основе базового формируется эталонное множество. Для этого базовое множество кластеризуется, и в эталонное множество включаются те слова, которые входят в кластеры, состоящие более чем из одного слова. Целью этого шага является исключение из базового множества тех слов, которые сочетаются с w лексически. Затем базовое множество пополняется теми словами, которые были отсеяны из-за низкой частоты $f(w, r, w')$, но которые при этом имеют достаточно малое расстояние до слов эталонного множества (расстояние вычисляется через контекстные вектора как описано выше). Как показали эксперименты, такое пополнение позволяет увеличить полноту базового множества более чем в 2 раза, не потеряв при этом точность. Пополненное множество снова кластеризуется, небольшие по размеру кластеры удаляются, а оставшиеся – используются для формирования описания семантической сочетаемости слова w .

Пятая глава посвящена проблеме интеграции словаря сочетаемости в систему автоматического синтаксического анализа. В разделе 5.1 предлагается метод, позволяющий учитывать содержащуюся в словаре сочетаемости информацию в процессе синтаксического анализа. Идея метода состоит в оценке вероятностей альтернативных синтаксических структур, возникающих в процессе анализа, исходя из сочетаемости слов, связанных в рамках этих структур синтаксическими связями. Данный метод был реализован и протестирован в рамках системы Treeton. Для этого в систему штрафных функций Treeton были добавлены средства оценки соответствия структуры сочетаемостным ограничениям, позволяющие взимать штраф за нарушение сочетаемостных ограничений, вычисляемый по следующей формуле:

$$penalty(Struct) = -\log \hat{P}(Struct) = -\log \prod_{w_1 \rightarrow_r w_2} \hat{P}(r, w_2 | w_1) = \\ \sum_{w_1 \rightarrow_r w_2} \left(-\log \hat{P}(r, w_2 | w_1) \right) = \sum_{w_1 \rightarrow_r w_2} \left(-\log \hat{P}(r | w_1) \hat{P}(w_2 | r, w_1) \right),$$

где произведение и сумма берутся по всем входящим в структуру *Struct* синтаксическим связям $w_1 \rightarrow_r w_2$. При вычислении $\hat{P}(r | w_1)$ используется информация о морфо-синтаксической сочетаемости слова w_1 , а при вычислении $\hat{P}(w_2 | r, w_1)$ – о его лексической и семантической сочетаемости. Таким образом, штраф синтаксической структуры сводится к сумме штрафов отдельных синтаксических связей, входящих в эту структуру (то есть обладает свойством аддитивности), что позволяет вычислять каждое его слагаемое только один раз – в момент построения соответствующей связи. Кроме того, функция *penalty* монотонно не убывает на частично-упорядоченном множестве синтаксических структур произвольного фиксированного предложения; при данных свойствах штрафной функции алгоритм синтаксического анализа гарантирует, что в произвольный момент времени все выданные результаты лучше (менее оштрафованы), чем все остальные. Это позволяет использовать первый выданный результат как наилучший, не дожидаясь окончания работы анализатора или даже прервав его работу, что чрезвычайно полезно в практических приложениях.

Экспериментальная проверка предложенного метода производилась на автоматически извлеченных из реальных текстов фразах, содержащих часто встречающийся тип неоднозначности – неоднозначность присоединения предложной группы (*смотреть соревнования по телевизору, смотреть соревнования по бадминтону*). Объем тестовых данных составил 3609 фраз, извлеченных из 30 текстов, не входящих в НКРЯ. Предложенный метод позволил улучшить точность разрешения неоднозначности данного типа с 54% до 89%. Также метод тестировался на составленных экспертами фразах, содержащих другие типы неоднозначности (202 фразы, 5 типов неоднозначности). В зависимости от типа неоднозначности точность варьировалась от 92% до 100%, средняя точность составила 94%.

В разделе 5.2 предложена технология формирования и сопровождения словаря сочетаемости и описаны поддерживающие данную технологию инструментальные средства, реализованные в рамках системы *Treeton*. Технология включает в себя следующие процессы:

- формирование базового словаря сочетаемости (занесение в словарь информации о сочетаемости часто употребляющихся слов, извлеченной из текстовой коллекции и доступных лингвистических источников автоматически);
- пополнение словаря в режиме обучения (включение в словарь автоматически извлеченной из текстовой коллекции информации о сочетаемости слов,

встретившихся в процессе синтаксического анализа; режим обучения используется при настройке на анализ текстов определенной предметной области, а также для развития базового словаря);

- тестирование системы синтаксического анализа и улучшение описаний сочетаемости (автоматическое сравнение результатов синтаксического анализа тестового корпуса с созданной экспертом эталонной разметкой; автоматизированное выявление и исправление ошибок анализа, вызванных неправильной информацией о сочетаемости).

Инструменты, разработанные для поддержки технологии, включают: инструменты формирования базового словаря на основе неразмеченной текстовой коллекции, базы данных моделей управления и толкового словаря; инструмент тестирования синтаксического анализатора; инструмент сопровождения словаря сочетаемости, позволяющий эксперту участвовать в процессе формирования описаний сочетаемости и исправлять возникающие в ходе данного процесса ошибки. При необходимости созданные инструменты можно с минимальными доработками использовать для работы с новыми лингвистическими источниками, имеющими схожую с поддерживаемыми источниками структуру (словари моделей управления, толковые словари), а также синтаксически размеченными текстовыми коллекциями.

В заключении сформулированы основные результаты диссертационной работы, обоснована их научная новизна и практическая значимость. Также предложены возможные направления дальнейших исследований.

Автор диссертационной работы выражает глубокую благодарность своему научному руководителю профессору М.Г. Мальковскому за внимание, помощь и поддержку на всех этапах выполнения работы, коллеге по научной работе А.С. Старостину за ценные советы и замечания, а также профессору Н.В. Перцову за чрезвычайно полезные консультации в области лингвистики.

ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ

1. Мальковский М.Г., Арефьев Н.В. Сочетаемостные ограничения в системе автоматического синтаксического анализа // Программные продукты и системы. №1. Тверь, 2012. – С.28-31.
2. Арефьев Н.В., Мальковский М.Г. Синтаксический анализатор Treevial. Оценка семантической корректности синтаксической структуры // Программные системы и инструменты: Тематический сборник. Под ред. Королева Л.Н. М.: Издательский отдел Факультета ВМиК МГУ; МАКС Пресс, 2009. № 10. – С. 100-107.
3. Арефьев Н.В., Булгаков И.А., Мальковский М.Г. Оценка качества и сопровождение синтаксического анализатора русскоязычных текстов //

Программные системы и инструменты: Тематический сборник. Под ред. Королева Л.Н. № 12. М.: Издательский отдел Факультета ВМиК МГУ; МАКС Пресс, 2011. – С. 111-124.

4. Арефьев Н.В. Использование онтологии для оценки семантической корректности синтаксической структуры // Сб. тезисов лучших дипломных работ 2008 года. М.: Изд-во факультета ВМиК МГУ им. М.В. Ломоносова, 2008. – С. 91–92.
5. Мальковский М.Г., Арефьев Н.В. Использование онтологии в системе автоматического синтаксического анализа // Сб. научных трудов по материалам международной научно-практической конференции «Современные проблемы и пути их решения в науке, транспорте, производстве и образовании '2008», т.2. Одесса: Черноморье, 2008. – С. 65-67.
6. Арефьев Н.В. Оценка достоверности синтаксических связей // Сб. научных трудов по материалам международной научно-практической конференции «Перспективные инновации в науке, образовании, производстве и транспорте '2009», т.2. Одесса: Черноморье, 2009. – С. 94-97.
7. Старостин А.С., Арефьев Н.В., Мальковский М.Г. Синтаксический анализатор «Treevial». Принцип динамического ранжирования гипотез // Компьютерная лингвистика и интеллектуальные технологии: по материалам ежегодной Международной конференции «Диалог» (Бекасово, 26-30 мая 2010 г.). Т. 9 (16). М.: Изд-во РГГУ, 2010. – С. 477-490.
8. Мальковский М.Г., Арефьев Н.В. Учет лексико-семантической информации в системе Treeton // Тихоновские чтения: Научная конференция, Москва, МГУ имени М.В. Ломоносова, 14 июня 2011 г.: Тезисы докладов. М.: МАКС Пресс, 2011. – С. 53.
9. Арефьев Н.В. Формирование словаря сочетаемости для системы автоматического синтаксического анализа // Сборник научных трудов SWorld. По материалам международной научно-практической конференции «Научные исследования и их практическое применение. Современное состояние и пути развития '2011», т. 4. Одесса: Черноморье, 2011. – С. 35-39.
10. Мальковский М.Г., Арефьев Н.В. Семантические ограничения в словаре сочетаемости: эксперименты по разрешению синтаксической неоднозначности // Сборник научных трудов SWorld. Материалы международной научно-практической конференции «Современные направления теоретических и прикладных исследований '2012». Выпуск 1. Том 4. Одесса: КУПРИЕНКО, 2012. – С. 21-25.

Напечатано с готового оригинал-макета

Издательство ООО “МАКС Пресс”

Лицензия ИД N 00510 от 01.12.99 г.

Подписано в печать 20.08.2012 г.

Формат 60x90 1/16. Усл.печ.л. 1.0 Тираж 70 экз. Заказ 314.

Тел. 939-3890. Тел./Факс 939-3891.

119992, ГСП-2, Москва, Ленинские горы, МГУ им. М.В. Ломоносова,
2-й учебный корпус, 527 к.